«Máquina de Ouver» — From Sound to Type

Finding the Visual Representation of Speech by Mapping Sound Features to Typographic Variables

João Couceiro e Castro
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
jccastro@student.dei.uc.pt

Pedro Martins
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
pjmm@dei.uc.pt

Ana Boavida
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
aboavida@dei.uc.pt

Penousal Machado
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
machado@dei.uc.pt

Figure 1: Representation of text before and after being run by the system.

ABSTRACT

Typography is considered by many authors the visual representation of language, and through writing, the human being found a way to register and share information. Throughout the history of typography, there were many authors that explored this connection between words and their sounds, trying to narrow the gap between what we say, how we hear it and how it should be read.

We introduce "Máquina de Ouver", a system that analyses speech recordings and creates a visual representation for its expressiveness, using typographic variables and composition.

Our system takes advantage of the scripting capabilities of Praat, Adobe's InDesign and Adobe's After Effects software to retrieve the sound features from speech recordings and dynamically creates a typographic composition that results in a static artifact as a poster or a dynamic one, such as a video.

The majority of our experimentation process uses poetry performances as the system input, since this can be one of the most dynamic and richest forms of speech in terms of expressiveness.

ARTECH 2019, October 23–25, 2019, Braga, Portugal
 ACM ISBN 978-1-450 3-7250-3/19/10… $15.00
https://doi.org/10.1145/3359852.3359892

CCS CONCEPTS

• **Information systems** → *Multimedia information systems; Information systems applications;* • **Applied Computing** → *Arts and humanities;*

KEYWORDS

Sound, Speech, Typography, Typographic Composition

ACM Reference format:

João Couceiro e Castro, Penousal Machado, Ana Boavida and Pedro Martins. 2019. «Máquina de Ouver» — From Sound to Type: Finding the Visual Representation of Speech by Mapping Sound Features to Typographic Variables. *In Proceedings of ARTECH 2019, 9th International Conference on Digital and Interactive Arts (ARTECH 2019),* October 23-25, 2019, Braga, Portugal. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3359852.3359892

1 INTRODUCTION

Since the dawn of mankind, oral communication has proven to be one of the most important skills for the survival and evolution of our species. With social and technological progress, the need to

preserve and share information in a physical way led to the invention of writing. What began as a three-dimensional system to keep track of goods using clay tokens as an accounting system, ultimately evolved to a two-dimensional abstraction and visual representation of sound through the conjugation of letters of the alphabet to represent phonemes [12].

The transition from calligraphic techniques to mechanical printing, in the 15th century, set the conditions to the mass production of printed materials but only in the early 20th century artists started to look at the printed page as an object of creative exploration [10].

In the works from the avant-garde poets and artists such as Stéphane Mallarmé (1842-1898), Filippo Marinetti (1876-1944), Guillaume Apollinaire (1880-1918), Hugo Ball (1886-1927), Ilia Zdanevich (1894-1975), Tristan Tzara (1896-1963) or Kurt Schwitters (1887-1948), it is possible to find some of the first exercises that explore the connection between the visual aspect of language and its sounds but the Robert Massin's graphic interpretation of Eugène Ionesco's "La Cantatrice Chauve" (1964) is one of the most interesting and relevant works, where typography is manipulated in terms of composition, white space, letter shape, weight, size, and slant in order to represent the actor's vocal expressiveness throughout the play.

With the rise of the computer era and all the contemporary digital tools to aid in the design process and development, the manipulation of typography and the exploration of its connection with sound became more accessible and, hence, present in the art and design scene. It is possible to see some results of that in the work of contemporary designers, such as Alan Kitching, Niklaus Troxler, David Carson, and Mitch Paone.

This research is focused on finding the best suited typographic variables that could represent measurable sound features in order to visually represent speech expressiveness in a systematic way.

In this paper, we start by presenting some related work from where we were able to take a few hints on the possibilities and limitations of our future project. Then, we present our "trial and error" based approach followed by an outline of our system, explaining each one of the process stages. For the experimentation section, we present four different phases that were crucial to the evolution of the system. All the conclusions taken from this research project so far, are summarized in the last section, as well as the expectations and suggestions for future work.

2 RELATED WORK

To the best of our knowledge, few are the research works that raise some questions and try to find answers on the topic of computational solutions to visual representation of speech expressiveness.

One of the first attempts to use software for exploring the temporal possibilities of digital media, in order to understand the communicational potential of kinetic typography, was made in 1997 by Ford, Forlizzi, and Ishizaki from Carnegie Mellon University [4] from which we can take some hints on how to manipulate typography in order to represent the linguistic sound features like pitch, loudness, and tempo. The use of size, position, weight, and color of the text is a clear option for that.

With an attempt to create a dynamic font based on the prosody of the speech, Rosenberg and MacNeil [11] identify some relevant results and conclusions like the visual changes made on the font being proportional to the intensity of the speech. A soft speech produces light weighted fonts with small sizes, while more intense speech produces bigger and heavier fonts. Another interesting aspect of this work is the possibility to apply changes at the word level or at the syllable level since the syllabic division makes the reading process closer to the way we hear oral speech.

The work developed by Fellows [3] tries to find the vocal tone of different movie characters and choose typefaces and styles that can represent them. Initial experiments try to make the connection between the main sound qualities, *i.e.*, pitch, intensity, duration, and timbre, and some typographic variables such as size, weight, and word and letter spacing.

In the work of Wofel, Schilipee, and Stitz [13], the authors' main focus is phonetics. Each grapheme design depends on the sound features extracted from its corresponding phoneme. Taking Paul Renner's Futura typeface as the base, the vertical and horizontal stroke weight, and character width are influenced by loudness, pitch, and speed.

3 APPROACH

Since the objective of this research work is to find a systematic solution to visually represent speech using typography as the main tool, as we can see in Figure 1, we propose that the solution to this problem resides in the definition of a set of rules, based on the mapping audio feature - typographic variable, to which we shall refer to as the "rule system".

The first approach to the attempt of finding this set of rules was to choose one audio recording and make an intuitive mapping. This was strictly based on what could be perceived as changes in intonation and expressiveness in speech.

In order to get a wide range of expressiveness, we decided to use recordings of poetry performances since this can be one of the most expressive forms of oral speech.

Not entirely satisfied with this experimental exercise due to the ambiguity and uncertainty of its results, the next step was to look into the field of Expressive Music Performance [2, 5, 7] and try to find some cues on which rules could be used to map sound to type.

Despite the fact that we are working with speech samples and not musical notes, the application of the rule system based on intensity and tempo led to some interesting results.

4 SYSTEM DESCRIPTION

As we can see in Figure 2, the pipeline of the system is composed of four stages: (1) Annotation, (2) Sound Analysis, (3) Data Treatment and (4) Artifact Generation.

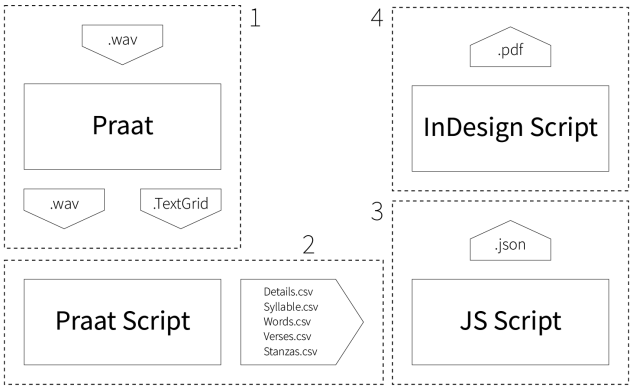

Figure 2: The pipeline of the system.

4.1 Annotation

To be able to dynamically produce the files containing the required data needed for the artifact generation, a prior annotation phase is imperative and requires the transcription of the sound recording.

The software used to annotate the verbal content of the sound file, as well as the later sound analysis, is Praat [1]. Being fairly easy to learn and use, well documented and with well-known results in the scientific community, Praat features a transcription tool and scripting capabilities, enabling the dynamic creation of files containing all the textual information and its corresponding measurement values and timestamps.

Praat annotation files, known as "TextGrids", can save both textual and time information and, in this case, must be created manually in the Praat application. TextGrid files can store information by layers called "Tiers" that can be of two types: interval tiers and point tiers. For more information about Praat Annotation feature visit www.fon.hum.uva.nl/praat/manual/. For this investigation, interval tiers are used to distinguish the different structural pieces of a poem, from the most general to the most detailed: stanzas, verses, words, syllables, and letters.

This method allows the next stages of the system to reach the full structure of the poem.

The TextGrid file that results from this stage, along with the Wave file, is used as input in the Sound Analysis stage.

4.1 Sound Analysis

The TextGrid file with the proper annotation and the Wave file with the speech recording are the inputs to a simple Praat script developed by the authors. This script takes advantage of Praat's Pitch and Intensity objects and the possibility of creating them dynamically from a given sound file with the Praat Scripting Language. Using these objects and the TextGrid file, this script is able to extract the start and end timestamps, duration, mean intensity and pitch values for every existing interval on each tier of the TextGrid.

This process results in a collection of Comma Separated Values (CSV) files, one for each tier containing all the intended data for each interval of the respective tier. These CSV files are processed in the Data Treatment phase.

4.2 Data Treatment

After the data acquisition from the sound file, there is a data treatment step where the CSV files are read, processed and merged into one single JavaScript Object Notation (JSON) file that represents the full poem.

The JavaScript script developed for this step serves two purposes: getting and writing the poem metadata and filter some possible errors that result from the Sound Analysis step.

Since there is the possibility of some pitch and intensity values being returned and written as "undefined", this script identifies them and rewrites them with the values resulting from a search for the next or previous interval corresponding value, depending on its existence.

The script then finds the minimum and maximum values for pitch, intensity, and duration and saves them as metadata inside the poem JSON object, making it the only input necessary for the Artifact Generation step.

4.2 Artifact Generation

The final step in the system chain is the Artifact Generation and the chosen developing environment was the Adobe's ExtendScript Toolkit with the help of the BasilJS library which makes scripting for Adobe's InDesign similar to Processing programming. For more on BasilJS, visit http://basiljs.ch/.

After creating a new Adobe InDesign Document, the script loads the poem JSON and maps the sound features' values to their respective typographic variables. This process is concluded by generating a PDF file with the final artifact.

5 EXPERIMENTATION

Being a "trial and error" approach, the experimentation process was heavily influenced by the subjective perspective of the authors and in this section, we present the leading questions, constraints, curiosities, and solutions to find a visual representation for speech expressiveness.

5.1 Intuitive Mapping

The first phase of experimentation was made without resorting to any computational techniques, entirely according to the intuition and perception of the authors.

5.1.1 Experimental Setup. The sound file used for this experimentation phase was the performance by the Portuguese actor and *diseur* João Villaret (1913-1961) of the poem "Cântico Negro", originally written by the Portuguese poet José Régio (1901-1969), retrieved from the album "João Villaret no São Luís" edited in 1961 by Valentim de Carvalho. The sound recording used in this test is available at http://bit.ly/2XhqtQ1. Being incredibly dynamic and expressive, this performance is a solid study object since it makes intensity, pitch and time variations easy to spot intuitively.

The only tool used to compose the typographic artifact for this attempt to map the sound to typography was Adobe's InDesign.

The typeface used to compose the artifact is TheSans, part of the Thesis superfamily, designed in the 1990s by the Dutch type designer Luc(as) de Groot.

Superfamilies consist of large collections of fonts with a considerable spectrum of alternatives that, since the 1990s, are designed and used to ease the process of font choosing and pairing. Sharing the base skeleton design, the combination of more than one font of the same family results in a balanced and harmonious composition.

In the scope of this experimentation, the choice of this typeface is centered on the various number of styles it offers and its neutral humanist design. Despite the numerous typographic classifications proposed over the years, some based on history and other purely on the design, we adopt the classification of American designer, writer, and educator, Ellen Lupton, and consider the humanist class as the roman typefaces from the 15th and 16th centuries that mimic classic calligraphy as of the contemporary typefaces that are inspired on them [8]. Based on that, these are considered the best ones to read in a large block of text.

Opting for the Thesis superfamily, setting the text with a large variety of weights and sizes becomes a possibility, without compromising its legibility, *i.e.*, the capability of differencing the glyphs while reading.

5.1.2 Experimental Results. The artifact that results from this experiment is partially presented in Figure 3 where the rules created during the process can be observed and identified.

For the extended pronunciation of letters, the visual solution used was the repetition of the characters that represent them. The size and weight of the font are used to represent the emphasis of the respective utterance, the more intense or high pitched the voice of the performer, bigger and heavier the font. Words spoken in a level of intensity that resembles screaming are capitalized.

For some characters, color is applied using a gray-scale palette to represent letters that are part of the word but aren't pronounced or, when used with the repetition of characters, to create a visual trail transmitting the idea of extending a letter with crescent intensity.

As for the positioning of the text on the page, there is no rule that specifies where it should be set. Regarding the positioning of a word relatively to its precedent, it is attempted that the visual horizontal space between them can be proportional to the pause duration in the speech. The same is also applied to the vertical space between verses in the form of leading, *i.e.*, space between lines of text.

5.1.3 Analysis. On further analysis of the resultant artifact, there are some notes worth register.

Font size and font weight are thought to be the two strongest typographic variables. Since there are no explicit features or their respective exact values being analyzed, the disparities in these variables are not fairly representative of the actual perceived differences in the speech.

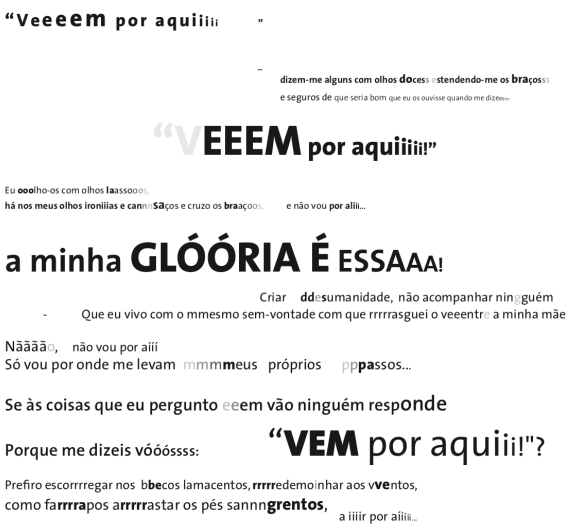

Figure 3: Intuitive Mapping Artifact Excerpt

The text is legible while the font size is not too low, therefore a minimum character size must be set. Being a non-serif typeface compromises it in terms of readability, *i.e.*, how fast a glyph can be perceived and, therefore, read.

A wide range of font weights proven to be a resourceful feature of the chosen typeface.

Using color can be a valid option but not a vital one since it is being used in a way that font weight or size can be used too.

The solution used for the pauses' duration during speech between words and verses is to be taken into account since it proves to be a natural use of the page's white space.

The ability to represent the duration of an utterance through the repetition of characters seems direct and intuitive but it raises questions on how to make the duration of a word visible and if the solution is anyway similar to the one used for the duration of the pauses.

Despite its high level of subjectivity and imprecision, the conception of this artifact proved to be a resourceful task allowing the authors to explore a premature set of rules and generate an artifact that encouraged them to engage in the next experimentation phase.

5.2 Expressive Music Performance Inspired Rule System

In this experimentation stage, we looked to the field of Expressive Music Performance, specifically to the work of authors like Gerhard Widmer, Werner Goebl, Elias Pampalk, Simon Dixon and Jorg Langner and try to mimic their approach of measuring the expression of a musical performance based on dynamics (intensity) and tempo [2, 5, 7].

5.2.1 Experimental Setup. For the first attempt to systematically map sound features to the respective typographic variables, the authors felt the need to simplify both the sound input and the number of pair rules between sound and type.

To use as sound input, the authors recorded several versions of the same sentence being said by the same person with different expressiveness levels as presented in Figure 4. By manipulating oral specifications like duration of the words, pauses, pitch, and intensity, it was possible to focus on how typography reacted to changes in sound. Sound recordings used in this test are available at http://bit.ly/2XhqtQ1.

The choice of sound features and typographic variables for these tests was centered in the use of intensity, and the duration of pauses and words for the sound, and size, weight and spacing for the characters.

The font used in artifact creation was changed from the previous experimentation phase from TheSans to the serif version of the same typeface Thesis, i.e., TheSerif.

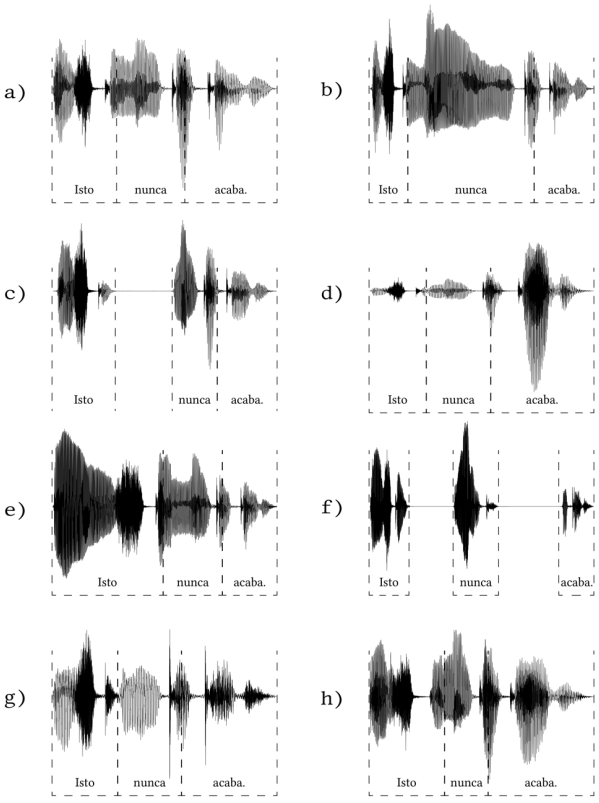

Figure 4: Sound Recordings' visual representation through their corresponding sound waves.

5.2.2 Experimental Results. In Figure 5 we can see the results of the first test in which the intensity of the voice varies between 30 and 82 and is mapped to the font size from 12 to 72 points.

The duration of each word is represented with tracking, i.e., the space between the letters of a word, varying from 0 to 500 and pause duration is represented by the space between words in the mean of kerning, from 0 to 5000.

From this experimentation was possible to identify that applying the rules on the word level results in imprecise artifacts and there are some dynamic changes that can be perceived in sound that are not represented. For example, observing Figure 4 is possible to see that every sound recording has a pause between the two syllables ``Is'' and ``to'' of the word ``Isto''. However, these pauses are so small, i.e., 0.03 seconds, that in further tests on the syllable level, resulted in an ambiguous visual representation. Based on the work of MacArthur, Zellou, and Miller [9] pauses under 0.1 seconds were ignored since they "do not consider pauses less than 100ms because fully continuous speech also naturally has such brief gaps in energy". Besides being possible to understand that tracking is representing the word duration, this does not seem the best option since the visual representation of slowly saying a word and the one from spelling it would be the same.

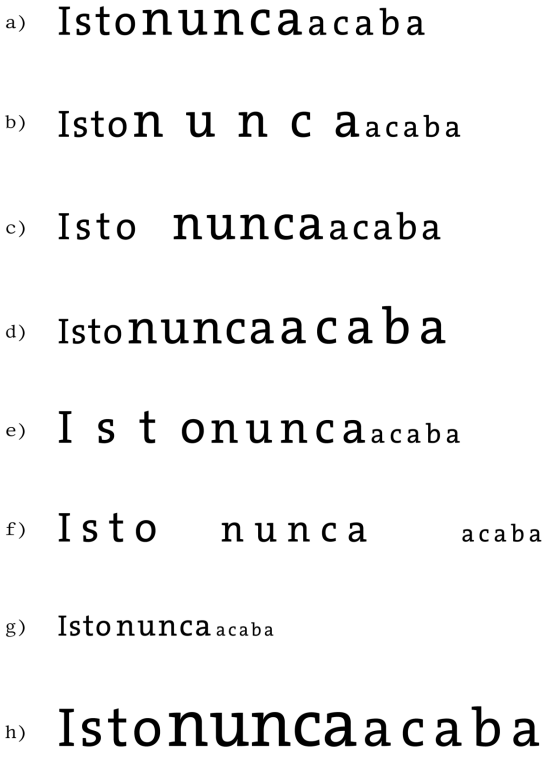

Figure 5: Results from the first automated test based on Expressive Music Performance.

For the next testing round, the rules were applied on the letter level, and the chosen visual cue to letter duration was based on a solution we all use every day while text messaging and mentioned in the work of Allan James [6] as an indication of "intensification of emotion or attitude" - character repetition. Hence, if the character duration is bigger than 0.2 seconds, it's repeated once for each 0.1 second interval. On example b) of Figure 6, it is possible to see how the 0.6 seconds duration of the letter "u" was mapped to a repetition of six characters.

Since font size and weight were identified as the most versatile and efficient typographic variables, tests were made where the

intensity was mapped not only to size but to font weight, using TheSerif's eight weight alternatives, from "Extra Light" to "Black".

As we can observe in Figure 6, both solutions result in much more dynamic artifacts than the previous tests and with a closer connection to the corresponding sound recordings.

During these tests, it was possible to identify a readability problem that resulted from the absence of space between words. Therefore, the existence of an empty space character was forced between every word and the impact of that decision can be observed by comparing the results from Figure 5 and Figure 6.

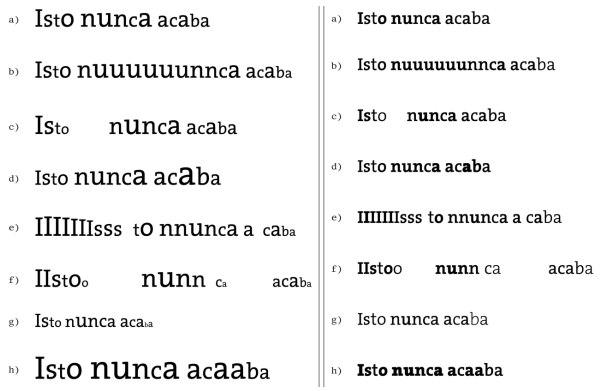

Figure 6: Results from mapping font size (on the left) and font weight (on the right) to the values of intensity.

5.2.3 Analysis. From the results of this experimentation phase, it was possible to confirm that both font weight and font size could be used to represent speech intensity and that this is a sound feature to include in the rule system. However, speaker voice pitch is another sound quality that is possible to extract and was not considered in this set of tests. In oral speech, the pitch variation is used, for example, as a way of giving the right intonation to a sentence so it can be perceived as an affirmation or a question. In the work of MacArthur, Zellou, and Miller [9] five of the twelve prosodic measures considered were based on pitch, so this is definitely a sound feature worth exploring.

Opting to apply the rules at the letter level and choosing to use the repetition of characters to represent the time speakers spend saying one particular letter, proved to be a valuable solution for our rule system, as well as the decision of forcing an empty space character between every word.

5.3 «Ouver» Rule System

Taking advantage of the conclusions from the first two attempts, in the third round of experimentation, the development of the system evolved towards a more complex set of rules and yet, a cleaner and more effective visual representation.

5.3.1 Experimental Setup. For this set of experiments, the input is an excerpt of "Amar ou Odiar", a poem written by the Portuguese author Fausto Guedes Teixeira, performed by João Villaret and retrieved from the album "João Villaret no S. Luís" edited in 1961 by Valentim de Carvalho. The sound recording used in this test is available at http://bit.ly/2XhqtQ1.

The objective of these tests is to apply the rules established so far in the previous rounds of experiments in a larger sound recording and to explore voice pitch as a valid sound feature, trying to find its typographic relative.

The typographic variables tested to represent voice pitch were position, color, font size, and weight.

5.3.2 Experimental Results. The first tests to find the visual variable for pitch representation were based in color and position.

On the left side of Figure 7 is the resultant artifact of mapping pitch variation to font size and baseline shifting, *i.e.*, the distance between the base of the letter and the imaginary line that all letters would normally sit on. On the right side of Figure 7, a gray-scale range of color is mapped to the same sound feature. In both examples, the intensity values are mapped to font weight.

In further tests, from which were extracted the artifacts present in Figure 8, font size and weight were alternately mapped to pitch and intensity in order to find the strongest option to shorten the link between speech and text. On the left side of Figure 8, the intensity was mapped to font size and the pitch to font weight. On the right side, the intensity was mapped to font weight and the voice pitch to the font size.

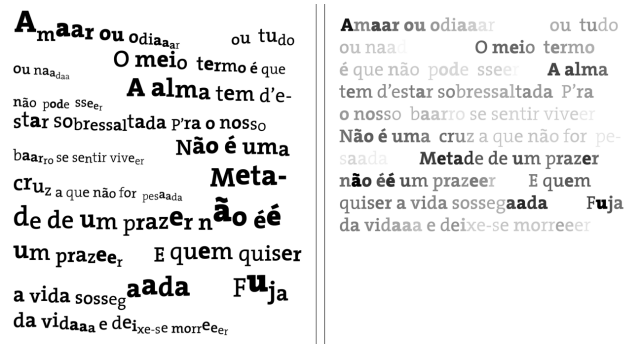

Figure 7: Results from mapping font size, and baseline shifting (on the left) and gray-scale color range (on the right) to values of pitch.

5.3.3 Analysis. The first tests to find the visual variable for pitch representation were based in color and position.

Despite the wide possibility range for the representation of pitch, font size and weight remain the main ones. By listening and reading along to the sound recording that produced the artifacts in Figure 8 the authors believe that assigning font weight to the speech intensity and font size to its pitch (right side of Figure 8 is the strongest option to achieve the best results.

In terms of text composition, since these artifacts are produced from poetry writings and performances, it would be interesting to test the generation of artifacts with the same textual structure as it was originally written but with the visual changes based on its performance.

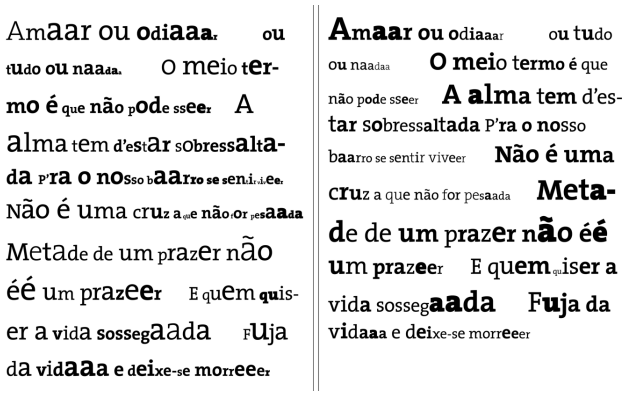

Figure 8: Results from mapping font size to intensity, and font weight to pitch (on the left) and font weight to intensity, and font size to pitch (on the right).

5.4 «Máquina de Ouver» Video Generator

In the final experimentation stage, a video generator was developed so it could be possible to dynamically reproduce the sound and the typographic manipulation in a synchronized way. 5.4.1 Experimental Setup. The sound input chosen for this stage was an excerpt from the 1980s' Portuguese television show called "A Dificuldade está na Escolha - Poesia Portuguesa I" where the Portuguese actor and declaimer Mário Viegas performs the poem "Cantiga dos Ais", originally written by the Portuguese poet Armindo Mendes de Carvalho. The sound recording used in this test is available at http://bit.ly/2XhqtQ1.

At this stage, the rule system is on its most recent version and the rules consist in mapping intensity to font weight and pitch to the font size. Since it is a poetic text, it's presented with the same textual structure that it was originally written and the pauses between words, syllables or letters are represented as horizontal white space and for the pauses between verses, the representation consists in the vertical white space.

For the video generation, the system exports an image for every change in text. Later, all the exported images are compiled with an Adobe's After Effects script that imports each image and sets its duration accordingly. The sound is imported, and the final video artifact is rendered.

5.4.2 Experimental Results. The resultant artifacts from this video generator represent one of the most interesting applications of the rule system since is possible to ear the changes in sound and follow them in text, as they are presented.

In Figure 9 is possible to see the text before and after the system processing and generation. The final video artifact is available at http://bit.ly/2XhqtQ1.

5.4.3 Analysis. In this stage we were able to obtain the dynamic effect of text visually reacting to sound, creating the possibility to follow the text and speech as it's said. Hence, video artifacts result in a more intuitive and perceptive experience than the static ones because they present the path between the first and more neutral form of the text to its last and unique representation.

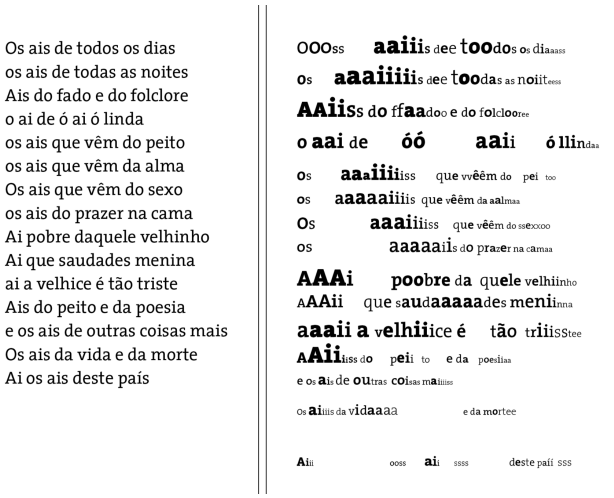

Figure 9: First and last frames of the video artifact generated by the system.

6 CONCLUSIONS / FUTURE WORK

From this work, we learned that it is possible to represent speech expressiveness using typography through the analysis of its sound features and structure.

At the current development stage of the system, the artifacts generated with the set of rules resultant from the experimentation process are promising and the link between sound and type is pleasantly noticeable. However, this approach and experimentation method relies mainly on the theoretical and empiric knowledge of its authors, being a subjective process. In order to test and validate the proposed solution, it's in authors best interest to elaborate a survey in which several groups of people with different backgrounds in typography, sound, music, reading, writing, and poetry could evaluate the poster and video artifacts generated by the system.

Being real-time speech expressiveness visualization one of the possible future outcomes of this research work, the manual transcription and annotation process is a significant limitation of this system. Future work may benefit from exploring the complementary implementation of speech-to-text and forced alignment tools.

With the results achieved in this research work, one future possibility is the development of a sound reactive variable font with self-mutating capabilities, that could be integrated into the system and used to generate the poster and video artifacts.

REFERENCES

[1] Paul Boersma and David Weenink. 2001. PRAAT, a system for doing phonetics by computer. *Glot international* 5, 341–345.
[2] Simon Dixon, Werner Goebl, and Gerhard Widmer. 2002. Real Time Tracking and Visualisation of Musical Expression. 2445.
[3] Kara S Fellows. (2009). *Typecast: the voice of typography*. Master's thesis. University of Iowa.
[4] Shannon Ford, Jodi Forlizzi, and Suguru Ishizaki. 1997. Kinetic Typography: Issues in time-based presentation of text. 269–270.
[5] Werner Goebl, Elias Pampalk, and Gerhard Widmer. 2004. Exploring Expressive Performance Trajectories: Six Famous Pianists Play Six Chopin Pieces.
[6] Allan James. (2017). Prosody and paralanguage in speech and the social media: The vocal and graphic realisation of affective meaning. *Linguistica* 57, 137.

[7] Jörg Langner and Werner Goebl. 2003. Visualizing Expressive Performance in Tempo—Loudness Space. *Computer Music Journal 27*, 4 (2003), 69–83.
[8] E. Lupton. (2010). *Thinking with Type, 2nd revised and expanded edition: A Critical Guide for Designers, Writers, Editors, & Students*. Princeton Architectural Press.
[9] Marit J MacArthur, Georgia Zellou, and Lee M Miller. 2018. Beyond Poet Voice: Sampling the (Non-) Performance Styles of 100 American Poets.
[10] P.B. Meggs, A.W. Purvis, and C. Knipel. 2009. *História do design gráfico*. Cosac Naify.
[11] Tara Rosenberger and Ronald L. MacNeil. 1999. Prosodic Font: Translating Speech into Graphics. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems (CHI EA '99)*. ACM, New York, NY, USA, 252–253.
[12] Denise Schmandt-Besserat. (2015). Writing, Evolution of. In *International Encyclopedia of the Social Behavioral Sciences (Second Edition)*. James D. Wright. Elsevier, Oxford, 761 – 766.
[13] Matthias Wölfel, Tim Schlippe, and Angelo Stitz. 2015. Voice Driven Type Design.